

G.P. Hammer<sup>1</sup>, J.B. du Prel<sup>2</sup>, M. Blettner<sup>1</sup>

# Vermeidung verzerrter Ergebnisse in Beobachtungsstudien – Teil 8 der Serie zur Bewertung wissenschaftlicher Publikationen

*Avoiding bias in observational studies – part 8 of a series on evaluation of scientific publications*

**Hintergrund:** Viele Fragestellungen im Gesundheitsbereich lassen sich nur mit Beobachtungsstudien untersuchen. Im Gegensatz zu kontrollierten Experimenten oder gut geplanten experimentellen, randomisierten klinischen Studien bergen sie einige Fallstricke, die zu verzerrten Ergebnissen führen können. Ein Grundverständnis dieser Probleme ist zur kritischen Würdigung entsprechender Publikationen notwendig.

**Methoden:** Hier werden einige wichtige Probleme von Beobachtungsstudien vorgestellt und mit Beispielen illustriert. Ergänzend wird auf selektierte Literatur verwiesen.

**Ergebnisse:** Faktoren, die zu verzerrten Studienergebnissen führen können, lassen sich grob einteilen in: Selektionsmechanismen bei der Rekrutierung der Probanden oder ihre kultur-, alters- sowie sozialstatusabhängige Teilnahmebereitschaft, uneinheitliche Datengewinnung, Messfehler, Confounder (Störgrößen) und weitere Fehler.

**Schlussfolgerungen:** Beobachtungsstudien leisten wichtige Beiträge zum Erkenntnisgewinn im Gesundheitsbereich. Wesentliche methodische Probleme lassen sich durch eine gute Studienplanung vermeiden. Die Kenntnis typischer Verzerrungsmöglichkeiten in Beobachtungsstudien ist bei der kritischen Lektüre von Publikationen notwendig. (Dtsch Zahnärztl Z 2011, 66: 212–217)

*Schlüsselwörter:* klinische Forschung, Studie, Beobachtungsstudie, Epidemiologie, Datenanalyse

**Background:** Many questions in human health research can only be answered with observational studies. In contrast to controlled experiments or well-planned, experimental randomized clinical trials, observational studies are subject to a number of potential problems that may bias their results.

**Methods:** Some of the more important problems affecting observational studies are described and illustrated by examples. Additional information is provided with reference to a selection of the literature.

**Results:** Factors that may bias the results of observational studies can be broadly categorized as: selection bias resulting from the way study subjects are recruited or from differing rates of study participation depending on the subjects' cultural background, age, or socioeconomic status, information bias, measurement error, confounders, and further factors.

**Conclusions:** Observational studies make an important contribution to medical knowledge. The main methodological problems can be avoided by careful study planning. An understanding of the potential pitfalls is important in order to critically assess relevant publications.

*Keywords:* clinical research, study, observational study, epidemiology, data analysis

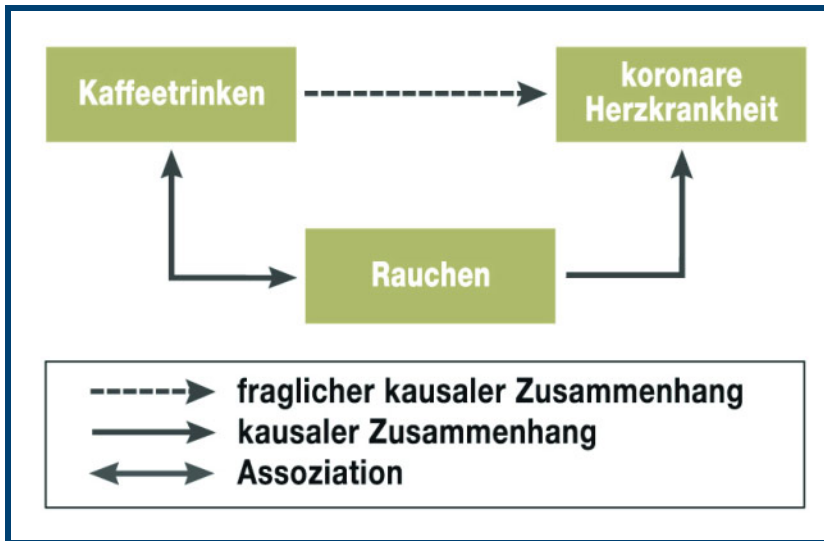
\* Nachdruck aus Dtsch Arztebl Int 2009; 106(41): 664–668; DOI: 10.3238/arztebl.2009.0664 © Deutscher Ärzte-Verlag GmbH Köln

<sup>1</sup> Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsklinikum Mainz

<sup>2</sup> Institut für Epidemiologie, Universität Ulm

Peer reviewed article: eingereicht: 17.10.2008, revidierte Fassung angenommen: 11.3.2009

DOI: 10.3238/dzz.2011.0212



**Grafik 1** Confounding: Rauchen, ein bekannter Risikofaktor für die koronare Herzkrankheit (hier der Endpunkt), der auch mit dem Kaffeetrinken assoziiert ist, täuscht einen kausalen Zusammenhang zwischen Kaffeetrinken und koronarer Herzkrankheit vor. Beispiel aus [2].

**Figure 1** Confounding: Smoking is a known risk factor for coronary heart disease – the endpoint here. As smoking is also associated with coffee drinking, this gives the false impression that coffee drinking and coronary heart disease are associated [2].

**Einleitung**

Die randomisierte klinische Studie ist in der klinischen Forschung ein fest etabliertes, häufig verwendetes Studiendesign und gilt gemeinhin als „Gold-Standard“ [1]. Viele Fragestellungen lassen sich aber nur mit epidemiologischen Beobachtungsstudien beantworten, wie zum Beispiel die Untersuchung des Einflusses von Zigarettenkonsum auf die Entstehung von Lungenkrebs [9], von Sport, Ernährung und Übergewicht auf Herz-Kreislauf-Erkrankungen [3] oder von UV-Exposition auf Hauterkrankungen [11]. Während in experimentellen, randomisierten klinischen Studien durch Randomisierung die gleiche Verteilung bekannter und unbekannter Störgrößen in den zu vergleichenden Gruppen erreicht werden soll, ist dies in Beobachtungsstudien selten möglich (siehe hierzu Teil 3 der Serie). Dies kann zu systematischen Verzerrungen und damit zu fehlerbehafteten Ergebnissen führen. Dieser Artikel soll aufzeigen, wie bei Studien, bei denen aus grundsätzlichen, ethischen Überlegungen keine Randomisation durchführbar ist, mögliche Fehlerquellen, die sich aus dem jeweiligen Studiendesign ergeben, erkannt werden können und wie man sie bei der Planung und Auswertung berücksichtigen kann.

Im Folgenden werden einige dieser Probleme beschrieben: Verzerrungen aufgrund von

- Selektionsmechanismen bei der Rekrutierung der Probanden (Selektions-Bias),
  - selektiver Erinnerung oder uneinheitlicher Datengewinnung (Informations-Bias), Messfehlern,
  - Confounding sowie
  - Simpsons Paradoxon und weitere Fehler.
- Ist man sich der Ursachen für Verzerrungen der Ergebnisse bewusst, können sie durch eine intelligente Studienplanung entweder ausgeschlossen oder reduziert werden. Zusätzlich sind diese Aspekte bei der Auswertung adäquat zu berücksichtigen. Dem kritischen Leser hilft ein Verständnis dieser Probleme bei der Interpretation von Studienergebnissen. Dem einführenden Charakter dieses Beitrags entsprechend werden Ergebnisse einer selektiven Literaturrecherche präsentiert.

**Ursachen von Verzerrungen, ihre Effekte und Gegenmaßnahmen**

**Selektionsbias**

Selektionsbias entsteht, wenn die Zielpopulation keine Zufallsauswahl aus

der Zielpopulation ist, für die eine Aussage getroffen werden soll. Probanden werden dann so rekrutiert, dass sie nicht repräsentativ für die Zielpopulation sind. Aber auch bei guter Planung kann es vorkommen, dass nicht alle ausgewählten Probanden an der Studie teilnehmen, allein schon deshalb, weil die Freiwilligkeit der Teilnahme immer gewährleistet sein muss.

In den folgenden drei Beispielen führt die Auswahl der Studienteilnehmer offensichtlich zu einer Selektion, die durch eine bessere Planung vermieden werden kann. Ähnliche Fehler werden leider immer wieder auch in Publikationen beobachtet.

– Das Gesundheitsamt einer großen Stadt möchte die Durchimpfungsrate der in der Stadt lebenden Vorschulkinder empirisch überprüfen. Dazu sollen die Impfpässe aller Kinder gesichtet werden. In drei Kindergärten machen die Eltern ausnahmslos mit, in den anderen Kindergärten ist die Teilnahme-rate gering. Ist das Ergebnis der Sondierung repräsentativ für alle Kinder? Wahrscheinlich nicht, denn es wurden nur Kinder aus bestimmten Kindergärten beziehungsweise Stadtgebieten untersucht. Kinder, die hierhin kommen, könnten sich in Merkmalen, die Einfluss auf die Impfbereitschaft der Familien haben, wie etwa dem Sozialstatus, von Kindern anderer Kindergärten unterscheiden. Die Bevölkerung, aus der die Probanden rekrutiert wurden, ist wahrscheinlich nicht repräsentativ für die Zielpopulation. Die Abhängigkeit der Durchimpfungsrate vom Sozialstatus ist bekannt [15].

– Mit einer anonymen Umfrage unter Forschungsnehmern hat das „US Office of Research Integrity“ sondiert, welcher Anteil der Wissenschaftler, die mit öffentlichen Geldern finanzierte Projekte durchführen, Forschungsergebnisse manipuliert hat. Dazu wurden die Probanden befragt, ob sie ein Fehlverhalten bei Kollegen beobachtet hätten [20]. Hier sind die durch ihr Teilnahmeverhalten selbst selektierten Probanden bestimmt nicht repräsentativ für die Zielpopulation aller geförderter Wissenschaftler.

– Ein Oberarzt möchte mehr über Risikofaktoren einer seltenen Erkrankung erfahren, für die er Spezialist ist. Seine Patienten nehmen weite Strecken auf sich, um sich an diesem Klinikum be-

handeln zu lassen. Dazu lässt er eine Doktorandin alle Erkrankungsfälle der letzten fünf Jahre erfragen und dazu (vom Alter und Geschlecht passende) Kontrollen aus dem Klinikum aus-suchen. Diese Kontrollen sind sehr wahrscheinlich nicht repräsentativ für die Bevölkerung, aus der sich die Fälle rekrutieren, denn sie kommen im Ge-gensatz zu den Fällen aus dem unmittelbaren Einzugsgebiet des Kranken-hauses.

Es ist schwierig, das Teilnahmeverhalten der Probanden zu beeinflussen. Ziel muss immer eine hohe Teilnahme-rate sein, um nach Möglichkeit einen reprä-sentativen Bevölkerungsquerschnitt zu erreichen. Auf jeden Fall muss in der Pu-blikation der Anteil von Nichtteilneh-mern angegeben werden. Meistens sind einige wenige Daten, wie zum Beispiel die Altersverteilung, über Nichtteilneh-mer bekannt. In vielen Studien wird au-ßerdem versucht, von Personen, die sich nicht an der Studie beteiligen wollen, zumindest eine kurze Auskunft, etwa in Form einer Postkarte mit wenigen Fra-gen, zu erhalten. Man fragt nach Grün-den für die Teilnahmeverweigerung. Diese Daten sollten bei der Interpretati-on der Ergebnisse berücksichtigt wer-den.

Eine Selbstselektion von Teilneh-mern findet auch statt, wenn Sprachbar-rieren oder gesundheitliche Barrieren die Teilnahme erschweren. Kulturelle Unterschiede und die soziale Schicht können sich auf die Teilnahmeberei-t-schaft zum Beispiel an Vorsorgeunter-suchungen auswirken. All das verringert die Verallgemeinerbarkeit.

**Informationsbias**

Informationsbias entsteht durch eine fehlerhafte oder ungenaue Erhebung in-dividueller Faktoren, seien es Risikofak-toren oder die untersuchte Erkrankung. Bei stetigen Größen (zum Beispiel Blut-druck) wird von Messfehlern gespro-chen, bei kategoriellen Merkmalen (zum Beispiel Tumorstadium) von Missklassi-fikation. Messfehler und Missklassifika-tionen entstehen seltener durch fehlen-de Sorgfalt der erhebenden Person oder mangelnde Qualität der Messgeräte/Er-hebungsinstrumente, als dadurch, wie und wann gemessen beziehungsweise klassifiziert wurde. Einige Fehler seien hier beispielhaft genannt:

Beispiel für Interaktion <sup>*1</sup>			
		Rauchen	
		nein	ja
Alkohol	nein	1,00 <sup>*2</sup>	1,53
	ja	1,23	5,71

**Tabelle 1** <sup>\*1</sup> Relatives Risiko von Alkoholkonsum und Rauchverhalten auf die Entwicklung von Mundhöhlenkrebs (aus [17]). Der Effekt des Alkoholkonsums ist bei Rauchern stärker als bei Nichtrauchern ausgeprägt.

<sup>\*2</sup> Die Bezugsbevölkerung ist die Gruppe derer, die weder rauchen noch Alkohol konsumieren. Ihr relatives Risiko ist per Definition 1,0.

**Table 1** <sup>\*1</sup> Relative risk of alcohol consumption and smoking on the development of carcinoma of the oral cavity (taken from [17]). The effect of alcohol consumption is greater in smokers than in non-smokers.

<sup>\*2</sup> The reference population in the group of individuals who neither smoke nor consume alcohol. Their relative risk is defined as 1.0.

	Therapie A (Therapieerfolge/ Patienten)	Therapie B (Therapieerfolge/ Patienten)
kleine Nierensteine	93 % (81/87)	87 % (234/270)
große Nierensteine	73 % (192/263)	69 % (55/80)
zusammen	78 % (273/350)	83 % (289/350)

**Tabelle 2** Beispiel zu Simpsons Paradoxon aus [6].

**Table 2** Example of Simpson’s paradox (taken from [6]).

- Typische Fragen zu weit zurückliegen-den Expositionen: In welchem Alter hatten Sie Windpocken und Masern? Wie viel Obst haben Sie letzte Woche verzehrt? Diese Fragen werden ver-mutlich mit einer großen Unschärfe beantwortet.
  - In einer Klinik werden morgens Blut-proben von Fällen und nachmittags von passenden Kontrollen genom-men. Anschließend werden sie zur gleichen Zeit nach einem einheitli-chen Verfahren analysiert. Leider be-wirkte die Dauer der Lagerung eine systematische Verzerrung (Bias) des Messergebnisses.
  - Mütter von Kindern mit Fehlbildun-gen erinnern sich besser an potenziel-le Risikofaktoren während der Schwangerschaft als andere Frauen (Erinnerungsbias) [21].
  - Ein Interviewer begegnet den befrag-ten Fällen mit mehr Empathie als den Kontrollen (da ihm der Status im Lau-fe des Interviews schnell bekannt wird). Dadurch bekommt er mehr und detailliertere Informationen von den Fällen (Interviewerbias).
- Die geschilderten Probleme kann man teilweise durch gute Planung umgehen, aber nicht immer durch die statistische Auswertung korrigieren. Interviewer-bias kann mit standardisierten Inter-views vermieden werden, in com-putergestützten Interviews können irre-llevante Fragen schneller übersprungen und widersprüchliche Angaben schnel-ler entdeckt werden. Der sensible Um-gang mit Tabus oder anderen kulturellen Unterschieden muss vor der Studie be-dacht oder gegebenenfalls getestet wer-den.

## Messfehler

Zusätzlich kann falsches oder ungenaues Messen zu Problemen führen. Systematische Messfehler entstehen zum Beispiel durch falsch geeichte Geräte. Zufällige, „klassische“ Messfehler kommen durch die Ungenauigkeit eines Geräts, eines Messverfahrens oder eines menschlichen Befunders zustande. Auch die nachträgliche Kategorisierung einer ursprünglich stetigen Variablen kann Messfehler nicht beseitigen und sollte vermieden werden [4].

Sind die Messfehler in ihrer Größe und Richtung bekannt, so kann man diese in der Auswertung berücksichtigen [16]. In einer Validierungsstudie müssen dazu (zum Beispiel bei einer kleinen Auswahl an Probanden) zusätzliche, genauere Messungen durchgeführt werden. In Ernährungsstudien wird beispielsweise oft das ungenauere Verfahren – der Fragebogen über den üblichen Verzehr – mit einem 24-Stunden-Verzehrprotokoll verglichen [16]. Die Beschreibung potenzieller Messfehler spricht für die Qualität einer Publikation.

In Bezug auf klassische Messfehler wird oft angegeben, dass sie die Studienergebnisse in Richtung eines „Null-Ergebnisses“ verzerren könnten. Nur unter sehr weitreichenden Annahmen kann man theoretische Überlegungen über Richtung und Größe der Verzerrungen machen. Diese Annahmen sind aber oft unrealistisch.

## Confounding

Unter einem Confounder versteht man einen Risikofaktor für die interessierende Erkrankung, der mit der interessierenden Exposition assoziiert ist und nicht in der Kausalkette zwischen der Exposition und dem Endpunkt steht (Grafik 1). Wird diese Assoziation im Untersuchungskollektiv bei der Auswertung nicht berücksichtigt – etwa, weil der Confounder nicht erhoben wurde –, führt das zu einer verzerrten Schätzung des Effekts des untersuchten Risikofaktors. Sind Risikofaktor und Confounder nicht miteinander assoziiert, wird der Effekt des Risikofaktors korrekt geschätzt.

Hierzu ein Beispiel für Confounding: Führt Kaffeetrinken zu koronarer Herzkrankheit? Das könnte man vermuten, wenn ein Zusammenhang beobachtet wurde [2]. Aber Kaffeetrinker sind

überdurchschnittlich oft Raucher, und es besteht neben der Korrelation zwischen Kaffee- und Nikotinkonsum ein starker kausaler Zusammenhang zwischen dem Rauchen und der Inzidenz der koronaren Herzkrankheit. Hier ist der Nikotinkonsum ein Confounder für den Effekt des Kaffeekonsums auf die Entstehung der koronaren Herzkrankheit.

Confounding ist dabei nicht zu verwechseln mit Interaktion („effect modification“): Zwei Risikofaktoren können entweder völlig unabhängig voneinander wirken, oder aber die Wirkung eines Risikofaktors hängt vom Vorliegen des anderen Risikofaktors ab.

Ein Beispiel für Interaktion: Aus Untersuchungen weiß man, dass sowohl Rauchen als auch Alkoholkonsum Risikofaktoren für die Entwicklung eines

### Wichtige Fragen bei Planung und Beurteilung von Beobachtungsstudien

- Ist das Studienkollektiv repräsentativ?
- Sind die betrachteten Teilkollektive vergleichbar?
- Sind die Informationen in vergleichbarer Weise erhoben worden?
- Werden potenzielle Messfehler beschrieben?
- Berücksichtigt das Studiendesign mögliche Fehlerquellen?
- Wie gut ist die Qualität der erhobenen Daten?
- Werden Korrekturverfahren angewendet?

Mundhöhlenkarzinoms sind. Die Risikoerhöhung durch Alkoholkonsum ist bei Rauchern stärker als bei Nichtrauchern ausgeprägt (Tab. 1) [17].

Confounding kann auf verschiedene Arten verringert werden. In klinischen Studien werden Patienten randomisiert auf die Behandlungsarme verteilt, in der Annahme, dass dann die Verteilung aller bekannten Störgrößen (wie Geschlecht und Alter) und sogar der unbekannteren Störgrößen in den Behandlungsarmen gleich sein wird (siehe auch Teil 2 der Serie). In rein beobachtenden Studien muss dagegen anders vorgegangen werden.

Eine Möglichkeit, den Effekt eines Confounders zu überprüfen, ist, das Untersuchungskollektiv in Schichten (Strata) zu unterteilen, die durch die Ausprä-

gungen des Confounders definiert sind. In unserem Beispiel könnten die Probanden stratifiziert werden in Nichtraucher, Probanden mit mäßigem Nikotinkonsum und starke Raucher. Die Auswertungen werden einmal unstratifiziert durchgeführt und einmal in den einzelnen Strata. Der Mantel-Haenszel-Schätzer [5, 14] wird oft dazu verwendet, die einzelnen Effektschätzer aus der stratifizierten Auswertung zu kombinieren und dabei den Confounder zu berücksichtigen. Je weniger sich die Ergebnisse der stratifizierten und der unstratifizierten Auswertung unterscheiden, desto geringer ist die Auswirkung des Confounders.

In Fall-Kontroll-Studien wird oft versucht, die Strukturgleichheit der Gruppen der Fälle und Kontrollen dadurch herzustellen, dass zu jedem Fall eine oder mehrere Kontrollen ausgewählt werden, deren Geschlecht, Alter und bekannte Confounder denen ihres Referenzfalls gleichen („matching“). Schüz und Kollegen untersuchten Risikofaktoren für Leukämie im Kindesalter und ordneten jedem Fall ein geschlechts- und altersgleiches Kind aus derselben Gemeinde zu [18].

Alle potenziellen Confounder beim Matching zu berücksichtigen ist selten möglich. Meist werden Beobachtungsstudien mit Regressionsmodellen ausgewertet. In diese Modelle werden die potenziellen Confounder – neben dem interessierenden Risikofaktor – als erklärende Variablen aufgenommen. Die Effekte der einzelnen Faktoren berechnet man dann adjustiert für die jeweils anderen. Den Effekt eines potenziellen Confounders kann man überprüfen, indem man die Ergebnisse aus zwei Modellen vergleicht, in denen er einmal in das Modell aufgenommen und einmal ausgelassen wurde. In Publikationen werden dann adjustierte und unadjustierte Ergebnisse nebeneinander präsentiert [13].

## Weitere Fehler

An dieser Stelle nennen die Autoren beispielhaft einige weitere potenzielle Fehlerquellen: Lead-time-Bias, ökologischer Trugschluss und Simpson-Paradoxon.

Nach der Einführung einer Vorsorgeuntersuchung werden meist längere Überlebenszeiten der Patienten beobachtet. Dies ist noch kein Beleg für den Erfolg

**Kernaussagen**

- Viele Fragestellungen zur menschlichen Gesundheit lassen sich nur mit Beobachtungsstudien beantworten. Wie jede Art von Studie sind sie potenziell mit Fehlern behaftet.
- Viele Faktoren können zu verzerrten Studienergebnissen führen. Sie lassen sich grob einteilen in Selektionsmechanismen, Messfehler, Störgrößen und methodische Fehler.
- Bestimmte Verzerrungsmöglichkeiten sind in Beobachtungsstudien aufgrund ihres Studiendesigns (zum Beispiel fehlende Möglichkeit der Randomisierung) häufiger. Die Autoren stellen wichtige Störgrößen vor und illustrieren diese mit Beispielen.
- Ist man sich der Ursachen für Verzerrungen der Ergebnisse bewusst, können sie durch eine intelligente Studienplanung entweder ausgeschlossen oder adäquat berücksichtigt werden.
- Dem kritischen Leser hilft ein Verständnis dieser Probleme bei der Interpretation von Studienergebnissen.

der Vorsorgeuntersuchung, denn die Patienten werden im Durchschnitt früher diagnostiziert und leben länger mit ihrer Diagnose. Dieses Phänomen ist als Lead-time-Bias bekannt. Es kann (teilweise) durch den Vergleich ähnlicher Regionen mit und ohne diese Vorsorgeuntersuchung und einer stadienspezifischen Auswertung berücksichtigt werden.

Eine Scheinkorrelation, wie zum Beispiel die von Höfer und Kollegen beschriebene Assoziation einer steigenden Anzahl Geburten, die außerhalb einer Klinik stattfanden, und einer parallel dazu steigenden Storchpopulation [12], verleitet dazu, einen Kausalzusammenhang zu vermuten, wo keiner ist. Fehler dieser Art können in ökologischen Studien auftreten, die ausschließlich aggregierte Daten auf Gruppenebene verwenden, zum Beispiel auf Gemeinde- oder Landesebene. Die beobachteten Zusammenhänge treffen aber nicht unbedingt auf die einzelnen Individuen der betrachteten Population zu. Die Kausalität

eines Zusammenhangs kann aufgrund fehlender Individualdaten nicht untersucht werden. Die Annahme der Übertragbarkeit von beobachteten Zusammenhängen von der Populationsebene auf die Ebene von Individuen nennt man ökologischen Bias oder ökologischen Trugschluss.

Weitere Scheinkorrelationen können entstehen, wenn Daten gruppiert ausgewertet werden, es aber innerhalb der Gruppen eine Ungleichverteilung einer wichtigen Größe gibt (die kein Confounder sein muss). Dieses Phänomen wurde unter dem Begriff Simpson-Paradoxon bekannt. Beispiele dafür finden sich auch in der Medizin:

In einem Vergleich zweier Therapien des Nierensteins [6] wurden die in Tabelle 2 aufgeführten Daten beobachtet. Wird die Größe der Nierensteine bei der Auswertung dieser Daten nicht berücksichtigt, scheint Therapie A eine schlechtere Wirkung zu haben (78 % versus 83 % Erfolg). Tatsächlich haben Patienten mit großen Nierensteinen eine schlechtere Prognose. Diese Patienten sind in Therapiegruppe A stärker vertreten. Dadurch ist der Behandlungserfolg von Therapie A scheinbar schlechter. Die Überlegenheit von Therapie A wird erst unter Berücksichtigung der Größe der Nierensteine ersichtlich.

**Fazit**

Beobachtungsstudien leisten wichtige Beiträge zur Kenntnis der Verteilung und der Ursachen von Krankheiten. Einige der Fallstricke, die zu verzerrten Ergebnissen führen können, haben die Autoren genannt. Beobachtungsstudien sind aber oft das einzige Mittel der Wahl, wenn es um lange Beobachtungszeiträume oder seltene Ereignisse geht oder wenn experimentelle Studien unethisch wären. Die wichtigsten Fragen bei der Planung und Beurteilung von Beobachtungsstudien sind im Kasten zusammengefasst.

Mögliche Fehlerquellen kann man oft mit Pilotstudien oder Prätests, also Voruntersuchungen mit der geplanten

Erhebungsmethode an einem kleineren Kollektiv, vor der eigentlichen Studierendurchführung rechtzeitig erkennen und korrigieren. Die Durchführung eines Prätests kann deshalb als ein Qualitätskriterium für eine Studie angesehen werden.

Wie viele Gedanken sich die Wissenschaftler gemacht haben, erkennt man an der Ausführlichkeit der Beschreibung der möglichen Schwächen einer Studie, der Methoden zur Vermeidung oder Korrektur absehbarer Probleme und des Umgangs mit unvorhergesehenen Gegebenheiten.

Als weiterführende Literatur sind die Leitlinie „Gute epidemiologische Praxis“ [8] und die Taschenbücher von Crombie und Greenhalgh zu empfehlen, die in Auszügen im British Medical Journal erschienen sind [7, 10]. Eine kurze, praktische Checkliste zur Bewertung (strahlen-)epidemiologischer Studien hat die Strahlenschutzkommission für eigene Zwecke erstellt und online publiziert [19].

**Interessenkonflikt:** Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

**Korrespondenzadressen**

Dr. P. H. Gaël P. Hammer  
Institut für Medizinische Biometrie,  
Epidemiologie und Informatik (IMBEI)  
Universitätsmedizin der Johannes  
Gutenberg-Universität  
Langenbeckstraße 1  
55101 Mainz  
E-Mail: gael.hammer@unimedizin-  
mainz.de

Prof. Dr. Maria Blettner  
Universitätsmedizin der Johannes  
Gutenberg-Universität Mainz  
Institut für Medizinische Biometrie,  
Epidemiologie und Informatik  
Langenbeckstraße 1  
55101 Mainz  
Tel.: 0 61 31 / 17 – 32 52  
Fax: 0 61 31 / 17 – 29 68  
E-Mail: maria.blettner@unimedizin-  
mainz.de  
www.imbei.uni-mainz.de

**Literatur**

1. Atkins D, Eccles M, Flottorp S et al.: Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. The GRADE Working Group. *BMC Health Serv Res* 4, 38 (2004)
2. Bonita R, Beaglehole R, Kjellström T: *Basic epidemiology*. 2nd ed. World Health Organisation, New York 2007
3. Boone-Heinonen J, Evenson KR, Taber DR, Gordon-Larsen P: Walking for prevention of cardiovascular disease in men and women: a systematic review of observational studies. *Obes Rev* 10, 204–217 (2009)
4. Brenner H, Blettner M: Misclassification bias arising from random error in exposure measurement: implications for dual measurement strategies. *Am J Epidemiol* 138, 453–461 (1993)
5. Breslow NE, Day NE: *Statistical methods in cancer research. Volume I. The analysis of case-control studies*. Lyon, France: International Agency for Research on Cancer 1980
6. Charig CR, Webb DR, Payne SR, Wickham JE: Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *Br Med J (Clin Res Ed)* 292, 879–882 (1986)
7. Crombie IK: The pocket guide to critical appraisal. BMJ Publishing Group, London 2004
8. Deutsche Gesellschaft für Epidemiologie (DGEpi) e.V.: Leitlinien für Gute Epidemiologische Praxis (GEP). <http://www.dgepi.de/pdf/infoboard/stellungnahme/GEP%20mit%20Ergaenzung%20GPS%20Stand%2029.7.2008.pdf>
9. Doll R, Peto R, Boreham J, Sutherland I: Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 328, 1519 (2004)
10. Greenhalgh T: *How to read a paper*. BMJ Publishing Group, London 2003
11. Hönigsmann H, Diepgen TL: UV-Hauttumoren. *J Dtsch Dermatol Ges* 3, 26–31 (2005)
12. Hofer T, Przyrembel H, Verleger S: New evidence for the theory of the stork. *Paediatr Perinat Epidemiol* 18, 88–92 (2004)
13. Korte JE, Brennan P, Henley SJ, Boffetta P: Dose-specific meta-analysis and sensitivity analysis of the relation between alcohol consumption and lung cancer risk. *Am J Epidemiol* 155, 496–506 (2002)
14. Kreienbrock L, Schach S: *Epidemiologische Methoden*. Spektrum Akademischer Verlag, Heidelberg, Berlin 1996
15. Morgenroth H, Hellenbrand W, Dreja I et al.: Die Durchimpfung von 24–30 Monate alten Kindern in pädiatrischen Praxen im Zeitraum von November 1999 bis Mai 2001 – Der Einfluss soziodemografischer Faktoren. *Gesundheitswesen* 67, 788–794 (2005)
16. Rosner B, Willett WC, Spiegelman D: Correction of logistic regression relative risk estimates and confidence intervals for systematic within person measurement error. *Stat Med* 8, 1051–1069 (1989)
17. Rothmann K, Keller A: The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *J Chron Dis* 25, 711–716 (1972)
18. Schüz J, Kaletsch U, Meinert R, Kaatsch P, Michaelis J: Risk of childhood leukemia and parental self-reported occupational exposure to chemicals, dusts, and fumes: results from pooled analyses of German population-based case-control studies. *Cancer Epidemiol Biomarkers Prev* 9, 835–838 (2000)
19. Strahlenschutzkommission: *Kriterien zur Bewertung strahlenepidemiologischer Studien*. Bonn: Strahlenschutzkommission 2002
20. Titus SL, Wells JA, Rhoades LJ: Repairing research integrity. *Nature* 453, 980–982 (2008)
21. Werler MM, Pober BR, Nelson K, Holmes LB: Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 129, 415–421 (1989)